



Interpreting Strokes on Paper with a Mobile Assistant

Theophanis Tsandilas

► To cite this version:

Theophanis Tsandilas. Interpreting Strokes on Paper with a Mobile Assistant. UIST - 25th Symposium on User Interface Software and Technology - 2012, Oct 2012, Cambridge, MA, United States. pp.299-308. hal-00722810

HAL Id: hal-00722810

<https://hal.science/hal-00722810>

Submitted on 4 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interpreting Strokes on Paper with a Mobile Assistant

Theophanis Tsandilas
INRIA - Univ Paris-Sud & CNRS (LRI)
F-91405 Orsay, France
Theophanis.Tsandilas@inria.fr

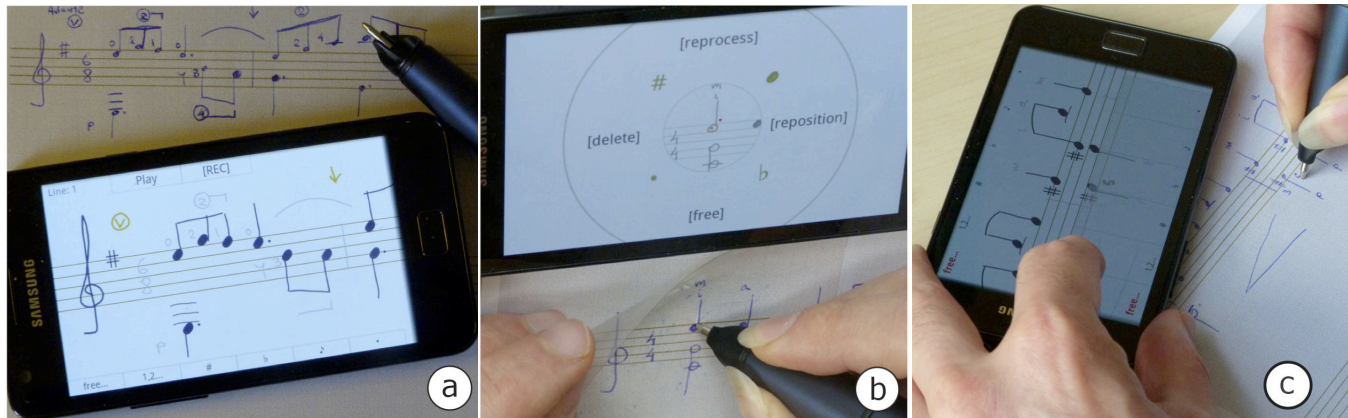


Figure 1. Writing music with an Aoto pen and a smartphone. (a) Handwritten score translated by the device. (b) Correcting the recognition of a note over a small plastic sheet. (c) Guiding the interpretation of strokes with the left hand.

ABSTRACT

Digital pen technology has allowed for the easy transfer of pen data from paper to the computer. However, linking handwritten content with the digital world remains a hard problem as it requires the translation of unstructured and highly personal vocabularies into structured ones that computers can easily understand and process. Automatic recognition can help to this direction, but as it is not always reliable, solutions require the active cooperation between users and recognition algorithms. This work examines the use of portable touch-screen devices in connection with pen and paper to help users direct and refine the interpretation of their strokes on paper. We explore four techniques of bimanual interaction that combine touch and pen-writing, where user attention is divided between the original strokes on paper and their interpretation by the electronic device. We demonstrate the techniques through a mobile interface for writing music that complements the automatic recognition with interactive user-driven interpretation. An experiment evaluates the four techniques and provides insights about their strengths and limitations.

ACM Classification: H.5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

T. Tsandilas. Interpreting Strokes on Paper with a Mobile Assistant. In *UIST '12: Proceedings of the 25th Symposium on User Interface Software and Technology*, ACM, 2012.

© ACM, 2012. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version will be published in *UIST '12*, October 7-10, 2012, Cambridge, MA, USA.

Keywords: Paper interfaces; bimanual interaction; pen + touch; mobile; interactive recognition; music interfaces.

INTRODUCTION

Previous studies suggest that paper is indispensable for several groups of users. Composers of contemporary music are proficient users of computers but still use paper at different stages of the composition process, from sketching initial ideas to working on final scores [11, 24]. Biology researchers use paper to record their field observations [26] or keep a history of their research protocols and results in their lab notebooks [19]. Digital pen technology has eased the connection between paper and computers. Yet, linking handwritten content with the digital world is still a challenging problem. The recognition of handwritten symbols is imperfect, while paper is not interactive by itself and the feedback provided by digital pens is still extremely limited.

Several approaches have succeeded in making paper interactive by using special equipment such as a projector and a tabletop [7] or a graphics tablet and a PDA [19]. More recently, Song et al. [22] explored paper interaction with a pen and a miniature projector. Here, we focus on the use of smartphones as mobile assistants for supporting work on paper and helping users direct the interpretation of their strokes. Smartphones are lightweight, inexpensive and widely available, and afford rich channels of input and output. Yet, how to take advantage of these additional channels when working on paper is a challenging problem that previous work has not examined. The paper explores techniques to help users interactively translate their writing on paper. We examine four forms of bimanual coordination (see Figure 2) that combine pen and finger touch and make

use of the smartphone's touchscreen to control how hand-written data are recognized by the computerized system.

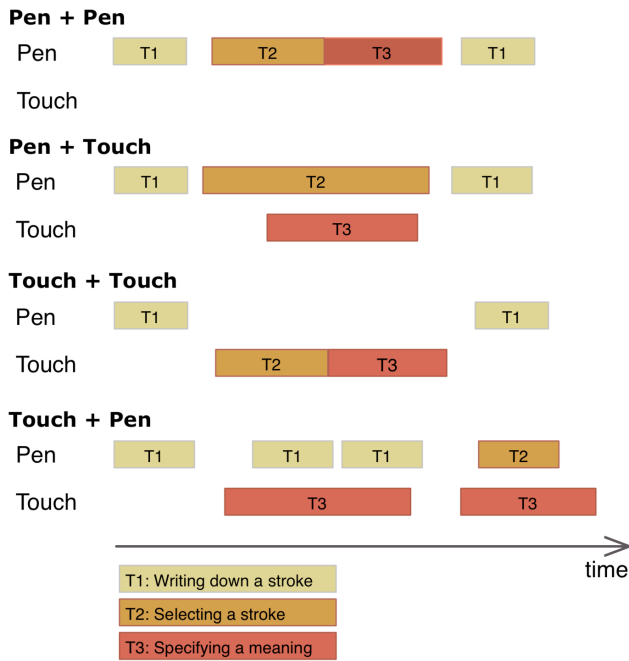


Figure 2. Forms of bimanual coordination. The pen is controlled by the dominant hand. Touch is controlled by the non-dominant hand.

To better understand how these techniques would be used in practice, we apply them to a concrete and challenging task, writing music (see Figure 1), which derives from real user needs [11, 24]. We demonstrate how they can enhance an interactive rule-based recognizer of musical notation by helping users correct recognition errors. We also show how the *Touch + Pen* technique can enable users to override the recognition process by constraining the interpretation of their strokes (see Figure 1(c)). Our approach has been based on Anoto paper technology, Bluetooth pens, and Android devices.

The paper presents an experiment that evaluates the four forms of bimanual coordination when recognition is imperfect. The results show significant performance benefits for the *Touch + Pen* technique, especially when the technique is used to constrain the interpretation of written strokes rather than correct recognition errors. User ratings, however, are not consistent with performance results, as *Pen + Pen* and *Pen + Touch* receive significantly higher scores. Based on our findings, we discuss the strengths and limitations of each technique and proceed with recommendations about their future use.

RELATED WORK

Interpreting Handwritten Data

Paper technology has been studied in range of application domains, from editing documents [17] to composing music [11, 24]. Interpreting the handwriting can be a difficult challenge in such domains, as data can take different repre-

sentations and forms that vary greatly among different users. Recognition algorithms can help to an extent, but their effectiveness is usually limited. For example, optical musical recognition for handwritten music has been available in commercial software. However, as Rebelo et al. [20] report in a recent survey, there are still no robust methodologies to recognize handwritten musical scores.

Researchers have explored a number of solutions to deal with the problem of recognition, but each solution has limitations. Some restrict the number of recognizable gestures and provide ink delimiters such as pigtails [17] and knotty gestures [25] to differentiate between commands and regular handwriting. Others, such as *Musink* [24], provide tools that help users refine the recognition but only after data have been uploaded to the computer. *Paper Substrates* [11] are specialized paper components that only accept certain types of data, which they know how to interpret and process. Finally, commercial products such as Livescribe provide buttons and menus printed on notebooks that activate explicit modes of stroke interpretation.

Pen, Paper and Mobile Feedback

Several solutions have studied the use of portable devices to assist interaction with paper although none of them deals with the problem of stroke recognition. Mackay et al. [19] introduce A-book, a portable device that helps biologists create links between their laboratory notebooks and online resources. A-book acts as a magic lens that displays digital information when positioned over a page. Yeh et al. [26] propose the use of a mobile device together with a digital pen to help biologists record data in the field. Finally, Liao et al. [18] use the camera of a mobile phone to make associations between digital and paper content. More recently, Heinrichs et al. [13] examine how users shift their focus between paper and a smartphone. They observe that users try to minimize the number of transitions between the two media, but transitions are not necessarily disruptive and often go unnoticed. Instead of a mobile phone, Song et al. [22] use a mini-projector to display information and menus on paper. Then, users interact with them by using the pen. Finally, Livescribe pens include a tiny OLED display and a built-in speaker to support live feedback.

Pen + Touch and Bimanual Coordination

Much of the HCI research on bimanual interaction builds upon Guiard's kinematic chain model [12] and its three main principles: (1) the dominant hand moves in the frame of reference defined by the non-dominant hand; (2) the two hands are asymmetric, the dominant one being more precise than the non-dominant one; and (3) the movement of the non-dominant hand precedes the movement of the dominant one. Yet, previous work [2] reports on asymmetric tasks that violate some of these principles. In addition, some results [23] suggest that the dominant hand may lead the non-dominant hand in symmetric tasks [23].

A number of approaches [7, 15, 27] explore forms of bimanual interaction for pen and touch and propose techniques of bimanual interaction that conform to Guiard's

model: select an object with the non-dominant hand and bring it under the pen, use the non-dominant hand to define a command mode or context that constrains the interaction with the pen, or make a rough selection with the dominant hand that is then refined by the pen. The pen is always held by the dominant hand, while touch is assigned to any hand. According to Brandl et al. [7], the combination of pen (dominant hand) with touch (non-dominant hand) is faster and more precise than using pens or touch by both hands.

In all the above approaches, the two hands move within the same reference frame and share the same views. Interaction with paper and a smartphone involves two separate spaces of both input and output. Balakrishnan and Hinckley [3] study bimanual interaction with separate reference frames but a common view. They show that users can effectively coordinate their hands as long as visual feedback is provided. In a later study of symmetric bimanual tasks [4], they observe that performance deteriorates when attention is divided between two visual spaces and propose that asymmetric tasks are more appropriate in this case.

BIMANUAL INTERPRETATION OF HANDWRITING

Our design approach is driven by phrasing [8] principles proposed by Hinckley et al. [15] for pen + touch interaction. Our goal was to avoid the use of persistent modes. We take advantage of the kinesthetic coordination of the two hands to augment the handwriting with computerized commands and user interaction. As a result, we do not require the use of any preprinted paper menus and palettes.

Scenarios

The need for data interpretation when working on paper applies to a range of real-world scenarios:

1. A student adds textual metadata, indexes and audio annotations to handwritten notes.
2. A musician works on a musical score on paper and produces its electronic version.
3. A researcher gets computational assistance while thinking over a set of mathematic equations.
4. A biologist records measurements in a laboratory notebook during an experiment.

In all these scenarios, a full or partial recognition of the handwritten strokes takes place. For example, the system must differentiate between commands, e.g., a command to activate an audio recording, and unrecognized text and identify tags that represent metadata. It must identify numerous symbols in a handwritten musical score, recognize their syntax and separate them from free annotations.

Strategies of Direct User Control

We identify two alternative strategies of user control that aid the interpretation of handwritten data:

- S1. The user relies on the system's ability to correctly recognize handwritten strokes and only intervenes to correct mistakes.
- S2. The user determines how the system should interpret pen strokes while he or she writes on paper.

The success of each strategy depends on how well the system recognizes the strokes and how easy for users it is to interactively direct or correct recognition. In addition, it depends on how well users could foresee the success or failure of the recognizer to optimize their strategies.

Forms of Bimanual Coordination

We studied the two strategies by focusing on three primitive tasks:

T1. Writing down a stroke or a group of strokes.

T2. Selecting one or multiple strokes.

T3. Specifying a meaning for the selected strokes or a context for their interpretation, e.g., a certain vocabulary that constrains their recognition. The user performs this task with a gesture or a menu selection.

The three primitive tasks can be carried out in various combinations, where pen and touch input can overlap in time. T1 is always performed on paper with the pen. T2 and T3 are performed by using either the pen or the touchscreen on the mobile device. We examined four alternative forms of bimanual coordination, presented in Figure 2. We name them after the sequence of the input channels involved in correcting recognition errors or specifying the interpretation of strokes. We assume that users always use their non-dominant hand to manipulate the mobile device, while their dominant hand holds the pen. All the four techniques allow for a posteriori error corrections (strategy S1). In addition to error correction, *Touch + Pen* can specify or constrain how the system will interpret a group of strokes (strategy S2). Past studies [9, 16] show that this last form of interaction is particularly effective for enforcing commands and switching between modes.

Interestingly, the four techniques do not all strictly obey the principles of Guiard's kinematic chain model [12]. In *Pen + Touch*, the dominant hand determines the reference of interaction and precedes the non-dominant hand, which makes the final selection. In comparison to how menus are activated and selected in the designs proposed by Brandl et al. [7], the technique inverses the use of the two hands. Notice, however, that precise task in our case is the stroke selection (T2). Respecting the second principle of Guiard's model, *Pen + Touch* assigns this task to the dominant hand.

APPLICATION: WORKING WITH MUSICAL NOTATION

We have designed and implemented a user interface for working with musical notation on paper that makes use of a digital pen and a smartphone (see Figure 1). Below, we present how the four techniques of bimanual coordination have been integrated into this interface.

Main User Interface

Figure 1(a) presents an example of a handwritten score and its translation by our system on a smartphone. Several strokes that represent key musical elements such as notes, stems, beams and symbols of intonation (e.g., sharps) have been recognized and replaced by dark glyphs. Other elements, such as rests and annotations, remain unrecognized;

these elements are displayed in a light grey color. Finally, some strokes that represent special functions, such as audio comments on the score, are shown in a yellowish color. The above interpretation has been made by an automatic online recognizer, but a set of bimanual interaction techniques has assisted the recognition process.

At every moment, the smartphone displays only a portion of the active staff. The user uses two fingers to scroll forwards or backwards along the active staff or simply taps the pen on paper to navigate within the score. When the focus of writing exits the active area of the screen, the screen scrolls with a fast animation, and the active stroke is centered at the first half of the screen.

How to represent the recognized symbols on screen turned to be a challenging design problem. The personal writing style of each user bears recognizable landmarks, essential for human recognition and navigation. Such landmarks are destroyed if the system replaces handwritten symbols by fonts that are commonly used in printed and electronic scores. Keeping or simply highlighting the original form of recognized strokes is not a satisfying solution either, because users need feedback about how the system interprets their handwriting. We took an intermediate approach. As shown in Figure 3-6, musical elements such as the head of notes, flats, sharps and keys are replaced by predefined beautified profiles and are positioned with respect to their recognized location in the score structure. Other elements such as stems and beams are slightly beautified but in a way that their relative geometry in the score is preserved.

Automatic Recognition

The user interface uses a custom online recognizer that supports the incremental recognition of the most basic elements of musical notation: notes (whole, half, quarter, eighth, etc.) and chords, ledger lines, dots, beams, flats, sharps, and common keys. Recognition takes place in two steps. Individual strokes are first classified by a tuned Rubine recognizer [21] into basic tokens: parallel and horizontal lines, curved lines or lines with a specific inclination, open and filled circles, special symbols such as G-clefs and flats. The result of this step and the original points of the stroke are then forwarded to a rule engine that keeps information about the changing structure of the score. The rule engine applies a set of syntax rules to derive the type of the new symbol and its connections with the existing score elements, recognized or not. If none of the rules applies, the stroke remains unrecognized. We have implemented a basic set of rules that capture common patterns of writing [1], but the recognizer can be extended with new rules.

The recognizer makes use of three types of rules:

1. *Recognition rules* that handle new strokes.
2. *Reinforcing rules* that reinforce the interpretation of a stroke or a group of neighboring strokes according to explicit user instructions.
3. *Overwriting rules* that handle the interpretation of strokes drawn over existing ones.

Reinforcing the Interpretation of Handwritten Strokes

The user can reinforce the recognition of strokes in a specific manner by using bezel menus around the screen with the left hand while using the right hand to write with the pen (*Touch + Pen*). We opted for a bezel menu as it affords tangible feedback and does not interfere with the content on the screen. Previous work has shown that bezel menus are particularly appropriate for eyes-free interaction [6, 16]. Our design has been inspired by the bezel menu proposed by Hinckley et al. [15] for creating new objects.

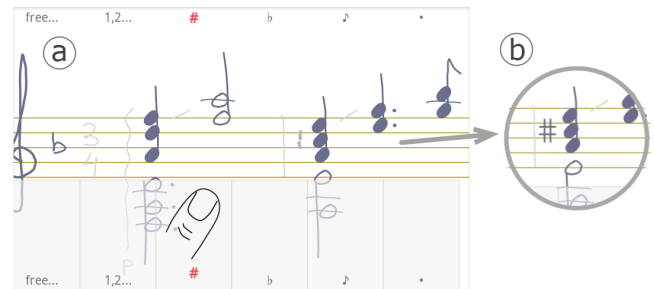


Figure 3. *Touch + Pen*: (a) Activation of the bezel menu at the bottom of the screen to reinforce the recognition of a sharp (#) while writing on paper (finger added only for illustration). (b) The recognized symbol after the entry of the strokes.

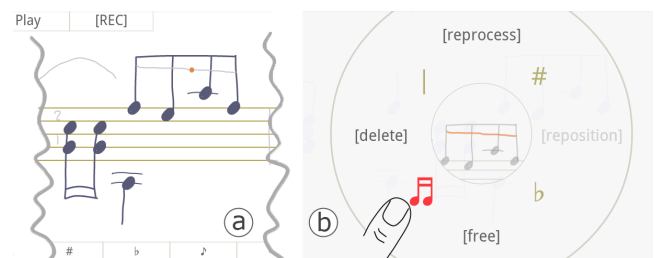


Figure 4. Using *Pen + Touch* to transform an unrecognized horizontal line into a beam: (a) Initial view (truncated). The user holds the pen over the line. (b) After a brief pause of 300 - 350 ms, a radial menu appears. The user selects the correct option with the finger while holding the pen on paper. A thumbnail of the score around the point of menu activation is shown at the center of the menu.

Figure 3 shows the user activating the bezel menu at the bottom of the screen by sliding her finger upwards. A semi-transparent layer follows the movement of the finger. The current selection (#) is shown in red. The menu options are also displayed at the top of the screen, relaxing the problem of hand occlusion as the finger manipulates the bezel menu.

Each option in the menu is associated with a single or multiple reinforcing rules. The rules are applied every time the user writes on paper while the corresponding option is active. For example, the second option ("1,2...") forces the recognition of strokes as numbers, and the 3rd option forces their interpretation as sharps. In contrast, the first option ("free...") deactivates the automatic recognition and handles strokes as free annotations.

Correcting Recognition Errors

The interface allows for correcting recognition errors. We have explored all of the four forms of bimanual coordination. The user can activate *Touch + Pen* with the bezel menu and then simply tap on individual or groups of strokes to reinforce their recognition. Figure 4 describes the error-correction mechanism for *Pen + Touch*. Holding the pen down over a handwritten stroke activates a radial menu at the center of the screen, enabling the user to assign a meaning from a contextual list of four alternatives or apply a fixed operation: *reprocess* a stroke, *reposition* it, treat it as a *free* annotation, or *delete* it. The menu appears when the pen pauses for 300 - 350 ms. This delay is short enough so that it does not interrupt the flow of interaction and long enough so that it is not accidentally activated when the user writes on paper. Subtle audio feedback informs the user about its activation.

The *Pen + Pen* technique provides a similar mechanism. The menu is activated in the same way, but now, the pen instead of the finger makes the final selection, with a marking gesture over the page. Since moving the pen on paper will leave an unwanted trace, the user has to activate the menu over a small transparent overlay that protects the page from inking (Figure 1). The role of the left hand is limited in moving and holding the overlay over the page. For the transparent overlay, we use pieces of plastic transparencies with low friction so that the pen does not leave any ink. The major strength of *Pen + Pen* is the fact that it relieves the user from interacting with the mobile device. The role of the device is now limited in providing feedback to the user. Here, we only examine the use of marking pen gestures to correct errors. However, we have also considered the use of quick symbolic gestures, such as the ones introduced by the Music Notepad [10].

To activate the radial menu with the *Touch + Touch* technique, the user has to point with the finger on the symbol of interest and perform a short slide movement. Figure 5 illustrates this mechanism. The sliding gesture is fast and relaxes the problem of occlusion: the ring provides a preview of the selection, allowing the user to return to the ring and cancel it before activating the menu.

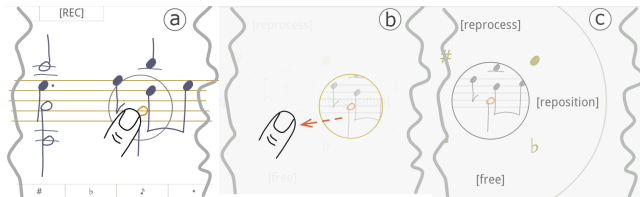


Figure 5. Activation of the radial menu with *Touch + Touch*: (a) A ring appears when the finger touches a symbol. (b) The finger slides out of the ring. (c) The finger is raised to activate the radial menu.

In addition to the above four techniques, we have implemented a simple correction mechanism, inspired from how users commonly make corrections on paper: the user re-writes over a symbol to replace it, hoping that the recognizer will make the correct decision. The mechanism is

associated with a set of *overwriting rules*. Each rule takes care of the replacement of a different type of symbol, such as the head of a note, a stem, or an unrecognized stroke.

Playing with Sound

The four techniques of bimanual coordination can support functions other than the recognition of musical notation. For example, the user can use the top bezel menu to play a segment of the score or create a link to a recording (Figure 6). To activate the first operation, the user must specify the start position by tapping the pen on the score while holding the *Play* item and release the finger to stop playing it. To create an audio link, the user must draw the linking strokes while holding the *REC* item and keep the finger down while recording. The user can later replay or remove the recording by selecting the strokes that represent the link with the finger (*Touch + Touch*) or the pen (*Pen + Pen* or *Pen + Touch*) and then select the target function (*Replay* or *Remove*) from a contextual menu.

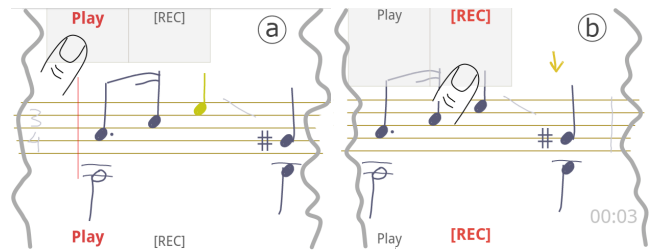


Figure 6. Using *Touch + Pen* (a) to play a segment of the score and (b) to create a link (yellow arrow) to an audio recording. The timer of the recording is shown at the bottom right of the screen.

Implementation Details

The implementation has been based on Anoto technology, streaming ADP-301 pens and Android smartphones. As we do not currently own pen drivers for the Android platform, we use a computer to mediate the communication between the pen and the smartphone. For Mac computers, we use the Bluetooth drivers of Letras [14]. In order to accelerate performance, preprocessing of the data with the Rubine algorithm is performed by a Java application on the computer. The rule engine, however, resides on the smartphone. As Android does not yet have support for midi, the computer is also responsible for reproducing the music. We use the JFugue API¹ to turn our internal score model into midi. The computer communicates with the device through OSC messages². Finally, we use the R-Tree structure as implemented by the SpatialIndex library³ to index strokes and accelerate their access from memory.

EXPERIMENT

We conducted an experiment to evaluate the four techniques of bimanual coordination on abstract tasks that do not require any domain expertise. The design of the techniques was very similar to one presented above. Our goals

¹ <http://www.jfugue.org/>

² <http://opensoundcontrol.org/>

³ <http://trac.gispython.org/spatialindex/>

were to test the performance of the techniques, explore how users would adapt them to their personal writing postures, and gather their personal preferences.

Participants

Sixteen volunteers (five women and eleven men), 25 to 39 years old (Mean = 30.9 years and Std. Dev. = 4.4), participated in the experiment. Two participants were left-handed. They all had experience with touch-based interfaces on mobile devices. Twelve of them were everyday users of smartphones or tablets.

Apparatus

Participants were seated in front of a desk. They interacted with an ADP-301 pen over seven preformatted A4 pages and a 5-inch 800×480 pixel Samsung Galaxy S. The connection between the pen and the Android device was mediated by a 2.66 GHz 15-inch MacBook Pro with 8 GB RAM. Pen data were received by the Android application in real-time with a latency of about 50 ms.

Task

Participants had to complete a series of writing tasks over horizontal lines printed on paper in a portrait orientation. Each writing task consisted of drawing strokes to replicate a group of three pre-printed symbols (see Figure 7). There were eight possible symbols:

| — ○ • × b ^ +

These symbols were presented in an arbitrary order, but the frequency of the first four symbols was twice the frequency of the rest. Handwritten symbols were drawn over the printed ones. Any stroke drawn away from the indicated symbols was ignored. Participants completed the tasks from left to right and from top to bottom. However, they could draw the three symbols of each task in any order. After finishing with the three symbols of a task, they had to draw a vertical *endline* to indicate the completion of the task. They could scroll along the timeline by dragging the screen with two or three fingers. However, this was generally not necessary as the screen was scrolled automatically whenever a stroke was drawn outside the screen.

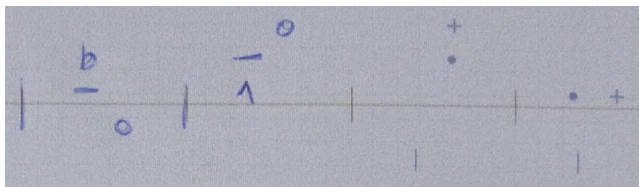


Figure 7. Extract from an experimental page. Symbols were positioned along several horizontal lines (experimental blocks) and organized by groups (tasks) of three symbols. The ending of each group was represented by a vertical *endline*.

Handwritten symbols were not always recognized by the system. The participant had to verify whether strokes were correctly interpreted by following their representation on the screen of the mobile device, shown in Figure 8(a). We expected that the accuracy of the handwriting recognition

would affect the performance of the techniques and yield a great variance between participants based on their writing style. Therefore, we decided to simulate and control the recognition instead of using a real recognizer. More specifically, recognition errors were triggered in an arbitrary but controlled manner: 25% of the tasks were error-free, 25% had one error, 25% had two errors, and 25% had three errors. We simplified the task by simulating a single type of errors: wrong negatives. This means that strokes were either correctly recognized or not recognized at all.

A drawback of our approach is the fact that recognition did not depend on how users wrote on paper. We tried, however, to eliminate any side effects of this problem by instructing the participants to write the correct symbols in a way that is proper for human readers. Participants were aware that they did not have any control of how the recognizer behaved. However, they were also informed that their pages would be later verified by a human.

Techniques

The experiment evaluated the four techniques of bimanual coordination discussed earlier: *Pen + Pen*, *Pen + Touch*, *Touch + Touch*, and *Touch + Pen*. Figure 8 illustrates how the interface on the device was adapted to the experimental task. For *Pen + Pen* and *Pen + Touch*, the delay for the detection of a pause and the activation of the radial menu was fixed to 320 ms. Participants used a transparent overlay of about 6×8 cm to interact with *Pen + Pen*. The bezel menu was only displayed for *Touch + Pen* and was positioned at the bottom of the screen. The size of its items was 88×35 pixels.

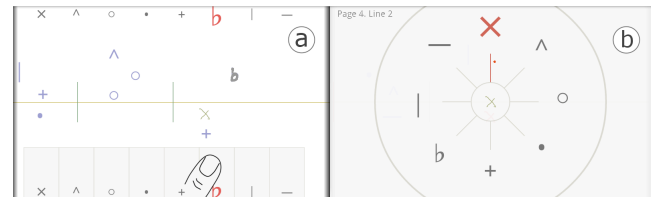


Figure 8. Representative screenshots from the mobile device. (a) The *Touch + Pen* technique over a line with recognized (bluish) and unrecognized (grey) strokes. Strokes are organized into groups of three, where each group represents a task. A vertical *endline* is drawn to complete a task. (b) Example of circular menu of eight items activated to correct a recognition mistake.

The strokes drawn on paper were mapped to the display with a scale of 10.12 pixels/mm. After the user activated the menu with the *Pen + Pen* technique, the scale was doubled, which allowed for menu selections with shorter marking gestures. Symbols on paper could be selected within a range of 0.51 mm around the rectangular boundaries of their strokes. This range corresponds to 5.2 pixels of screen size. Our pre-tests showed that this range was not enough for the *Touch + Touch* technique. Therefore, for this technique, we set the range of touch selection beyond the boundaries of the strokes to 10 pixels.

Conditions

The techniques were tested under two conditions that capture the two extremes of real conditions:

Unpredictable. Participants had no information about whether the system would succeed in recognizing a stroke. Participants had to depend on the visual feedback of the mobile device to anticipate an error.

Predictable. Participants knew beforehand whether recognition would succeed or not. Symbols corresponding to recognition errors were printed in red, which allowed the participant to develop a strategy before starting a task. This condition simulates the situation in which the user writes a special symbol, knowing that the symbol does not belong to the recognizer's vocabulary.

Design and Procedure

We followed a mixed full-factorial design: *Condition* was treated as a between-participants factor whereas *Technique* was treated as a within-participants factor. For each technique, participants completed one page of six lines with eight tasks each. The design can be summarized as follows:

- 2 Conditions (*Unpredictable*, *Predictable*)
- × 8 Participants per condition
- × 4 Techniques
- × 6 Lines (blocks)
- × 4 Error Levels (0, 1, 2, and 3 errors)
- × 2 Repetitions (each with a different set of symbols)
- = 3072 tasks in total

In addition to the main tasks, participants completed 2- 3 lines of practice tasks per technique on three separate sheets. The order of presentation of the four techniques was balanced among participants with a Latin square. The position of the items in the menus changed for each page but remained the same for all the tasks on the same page.

We asked participants to complete each task as fast as possible, trying to write clearly for humans and avoid mistakes. Participants could take pauses between tasks. They were encouraged to adjust their posture and re-position the page and the mobile device during the tasks to optimize their comfort and performance. Every time a new page was given to a participant, the experimenter placed the device at the center of the page, which required the participant to move it around. The experimenter also provided a cloth that could be placed under the device to increase the friction with the table and prevent accidental displacements while touching on the screen. The experimenter took notes during the sessions about positioning of the mobile device. After the end of the session, participants completed a questionnaire with 5-level Likert-scale questions about their experience with the four techniques. The whole procedure lasted for approximately 50 - 70 minutes.

Measures and Hypotheses

We measured the time needed to complete a task. It was measured from the point the participant started drawing the first symbol of the task (or from the point that the partici-

pant started touching the screen for the *Touch + Pen* technique) to the point the participant finished drawing the vertical end-line. We also counted the number of errors, i.e., untreated recognition mistakes or wrong assignments to symbols, as well as pointing errors occurring during the task, i.e., failure to select a stroke by touch or pen. Finally, we collected participants' responses concerning various qualities of the tested techniques: learning time, performance, precision, physical demand, attention, suitability for real applications and overall preferences.

The experiment was mostly exploratory since we did not have clear hypotheses about the relative performance of the four techniques. Based on past results [7], we expected that the pen would be more accurate than touch in making precise selections. We also expected that the predictability factor would have a clear effect on user performance. The *Unpredictable* condition introduced an additional cost, as participants had to pay continuous attention to the screen of the smartphone, keeping track of how the system interpreted their strokes. We predicted that this effect would be more intense for *Touch + Pen*, as this technique required the user to choose between the two possible strategies: correcting recognition errors (S1) or specifying the interpretation of strokes while writing on paper (S2).

Results on User Performance

We first analyzed task completion times. A total of 154 tasks or 5% of time measurements were disregarded from the analysis due to missing data, errors (2.8% of tasks) and outliers (1.9% of tasks) three standard deviations away from the within-cell mean. The overall results for task completion times are shown in Figure 9.

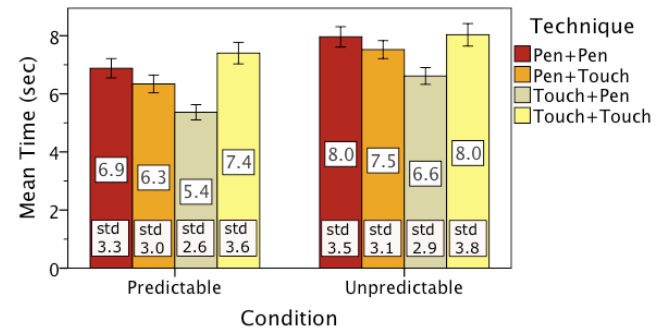


Figure 9. Results for task completion time. Error bars represent 95% Confidence Intervals (CI). Standard deviations are shown at the bottom.

We conducted an ANOVA analysis with *Technique*, *Line*, *Error Level* and *Repetition* treated as repeated-measures factors and *Condition* treated as a between-participants factor. The main effect of technique was statistically significant ($F_{3,42} = 20.6$, $p = 2.3 \cdot 10^{-8}$). A pairwise comparison with Bonferroni's adjustment for six pairs of comparisons showed that *Touch + Pen* was statistically faster than any of the other three techniques ($p < .01$). However, we found no significant difference between any other pair. The difference between *Pen + Touch* and *Touch + Touch* was only marginally significant ($p = .086$).

As we expected, the main effect of *Condition* was statistically significant ($F_{1,14}=5.04$, $p=.041$). The completion time under the *Predictable* condition ($M=6.5$ sec, $SD=3.3$ sec) was about 14% faster than the time under the *Unpredictable* condition ($M=7.5$ sec, $SD=3.4$ sec). Interestingly, the mean difference between the two conditions for tasks with no recognition errors (*Error Level* = 0) was 1.3 sec, which translates into a 47.7% increase in time. This difference represents the extra cost due to divided attention between the smartphone and the page as participants tried to keep track of unpredictable errors.

We found no significant interaction effect between *Technique* and *Condition* ($F_{3,42}=.697$, $p=.56$). Not surprisingly, the main effect of *Error Level* was significant ($F_{3,42}=704.8$, $p=6.0 \cdot 10^{-36}$), with time increasing linearly ($F_{1,14}=933.4$, $p=3.2 \cdot 10^{-14}$) in relation to the number of recognition errors per task. We also found an interaction effect between *Error Level* and *Technique* ($F_{9,126}=15.7$, $p=6.2 \cdot 10^{-17}$). Figure 10 presents the effect of *Error Level* for each technique and summarizes the results for the pairwise comparisons.

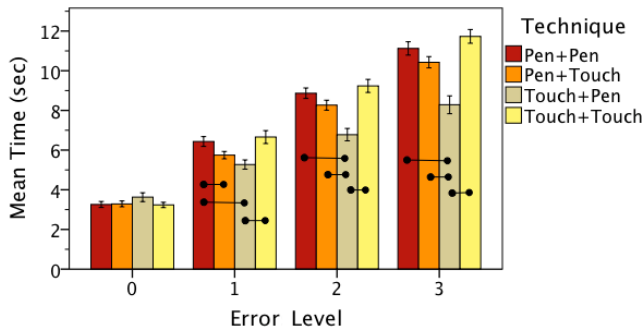


Figure 10. Task completion time per *Error Level* and *Technique*. Error bars represent 95% CI. Connections between bars represent statistically significant differences ($p < .05$), after Bonferroni's adjustment for 24 pairwise comparisons.

We examined the strategy that participants followed for *Touch + Pen*. Four participants used the bezel menu to only correct mistakes (strategy S1); seven participants used it to assign symbols when drawing strokes (strategy S2); and five participants mixed both strategies. There was no clear relationship between strategy and condition. As shown in Figure 11, S2 seems to be the most effective strategy.

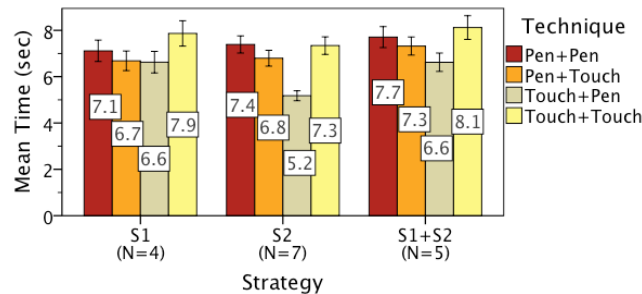


Figure 11. Results by participant group. Groups are based on selected strategies for *Touch + Pen*. Error bars represent 95% CI.

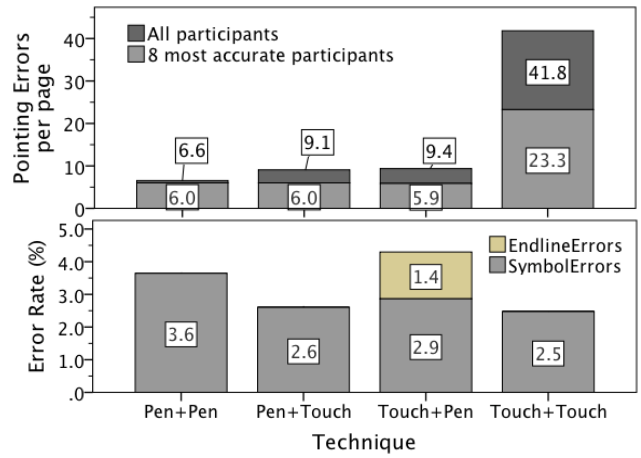


Figure 12. Distribution of errors across techniques

We also analyzed the errors made by participants. Figure 12 (bottom) presents the mean number of untreated recognition errors per task. It also shows the additional errors caused by *Touch + Pen*, when participants kept holding an item of the bezel menu while completing a task. Now, if we examine the error rates with respect to the strategy followed by participants for *Touch + Pen*, we can observe an interesting asymmetry: 0.5% for the four participants that chose the slow strategy (S1), 4.7% for the seven participants that chose the fast strategy (S2), and 1.7% for the five participants that combined both strategies.

The top chart in Figure 12 shows the distribution of pointing errors when participants tried to correct a symbol. Clearly, touch was less precise than the pen. Friedman's non-parametric test showed that the effect of *Technique* on pointing errors was statistically significant ($\chi^2(3)=30.66$, $p<.001$), while pairwise comparisons showed that *Touch + Touch* produced significantly more pointing errors (adjusted $p<.001$) than any of the three other techniques.

Positioning the Device in the Physical Space

We observed a variety of strategies concerning the manipulation and positioning of the smartphone. Eight participants were very active in moving the device during the tasks to minimize the distance between the screen and the writing zone and improve the coordination of their hands. Three more participants moved the device but less frequently or only for certain techniques. Finally, five participants were passive, initializing the position of the device before starting writing on a new page and keeping this position for all the tasks. Several of these participants, however, frequently changed the posture of their body and kept moving the page to align the active zone of writing with the device.

Subjective Ratings and User Suggestions

Figure 13 presents participants' subjective ratings. Surprisingly, the ratings are not consistent with performance results. *Pen + Pen* and *Pen + Touch* were rated the highest across all variables. *Touch + Pen* received low ratings even with respect to speed. It seems that participants underestimated the time required to correct recognition errors and the ability of the technique in saving time. Some partici-

pants expressed the opinion that *Touch + Pen* required additional concentration and effort because they had to synchronize both hands at the same time and to frequently shift their attention from the page to the device. According to a participant, the technique "*was confusing and [I] always had to keep an eye on the touchpad to make sure I had the correct item selected*".

We received a number of suggestions about how to improve the design of the techniques. Several participants complained about the occlusion of the finger on the touch-screen and the high precision required to select strokes. Some participants found that the delay (320 ms) required to activate the menu with the *Pen + Pen* technique was too long, as they were tempted to start the marking gesture immediately after putting the pen on the transparent overlay. Two participants suggested that *Pen + Touch* must be enhanced with an additional menu-activation mechanism: the radial menu appears when the user pauses the pen when drawing and disappears if the movement restarts. Another participant asked for richer audio feedback to (1) avoid touch errors when the attention was on the page, and (2) avoid pen errors when the attention was on the screen.

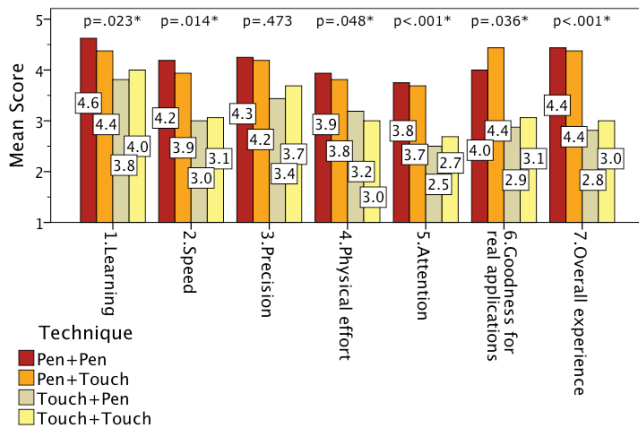


Figure 13. User ratings about various usability variables in a Likert scale (5 = Positive, 1 = Negative). The p values above the bars give the results of Friedman's non-parametric statistical tests for the effect of *Technique* on the score of each variable.

DISCUSSION

Touch + Pen was clearly the fastest technique. This result was independent of the condition, i.e., whether recognition errors were predictable or not. It seems, however, that there was a correlation between the performance of *Touch + Pen* and the strategy that participants followed to complete the tasks. Of course, given the limited number of participants, we cannot make safe conclusions. But our results suggest that participants who consistently used strategy S2 managed to gain a 24 to 30% in speed in comparison to the three other techniques. On the other hand, strategy S2 seems to inflate errors, which may explain why four participants did not use it all and why five other participants mixed it with strategy S1. Reflecting on the use of strategies in real scenarios, a participant commented:

"I would use the technique with both strategies: first, a lot the second one, then I would progressively switch to the first one, provided that the recognizer improves."

Given now the fact that *Touch + Pen* was rated low by most participants, we can conclude that the technique must be used sparingly, in situations in which its use yields significant benefits. It could be especially appropriate for symbols that appear repetitively and in sequence, such as multiple sharps (#) at the beginning of a score, and whose recognition is difficult and imprecise, such as newly added symbols for which the recognizer has not yet training samples. Also, it can be particularly useful for assigning meaning to multi-strokes and groups of symbols, as it is the only technique that can delimit a sequence of pen actions without requiring the use of persistent modes. The fact that *Pen + Touch* was highly preferred to *Touch + Pen* by participants implies that Guiard's principle about the sequential order of the two hands may have exceptions, at least when hands operate in different reference frames, the views are split, and precision is critical for both hands (e.g., making selections in the bezel menu with the non-dominant hand), which means that tasks are not entirely asymmetric.

Pen + Pen and *Pen + Touch* had a similar performance with differences in speed less than 10%. *Pen + Touch* was shown to be significantly faster only for tasks with a single error. In addition, the two techniques received similar ratings by participants. Each technique has different advantages and weaknesses. Fortunately, one technique does not exclude the other, and users can make use of any of the two depending on their work strategies. *Pen + Pen* does not require users to directly interact with the smartphone. They can use it as a peripheral device that supports awareness and only shift their attention to it when they correct an error. A drawback of the current implementation of the technique is the fact that the pen cannot detect the overlay and automatically switch to a command mode. A pause is necessary, which breaks the flow of interaction. As a participant of a pilot study commented, the technique requires the user to mentally switch states that do not directly correspond to the visual and tangible feedback that the interaction instruments [5] provide. Garcia et al. [11] explore the use of active transparent layers detected by the pen, but their mechanism involves a time-consuming calibration phase. We expect, however, that technology of digital pens will soon resolve this limitation.

The *Touch + Touch* technique proved to be the least accurate despite the fact that its stroke selection was given some small advantage compared to the other techniques. We expect that this problem will become more intense in smaller devices. Finger occlusion is not the only issue. The use of the non-dominant hand in conjunction with the varying parallax caused by moving the device around the page aggravates the problem. A participant explained that the technique would be his favorite if the problem of precision were somehow resolved. Consider that *Touch + Touch* is the only technique that makes a clear separation between

writing and error correction. When working with musical notation, precision is important as symbols are tiny, are stacked together, and appear in close proximity. The technique, however, could be useful for the selection of larger entities, such as chords and measures, or special elements such as links. Finally, the technique can be valuable in other types of tasks, such as adding tags to text, where interactive elements are sparse, so precision is not an issue.

SUMMARY AND FUTURE DIRECTIONS

We examined how mobile assistants can help users interactively guide the interpretation of their handwriting when working on paper. We identified four techniques of bimanual coordination that serve two strategies of direct user control: (1) correcting recognition errors, and (2) constraining recognition while writing. We explored the design of these techniques with a user interface that helps users work with classical musical notation. An experiment evaluated the four techniques. Its results provide valuable feedback about the performance of the techniques, but also about their strengths and limitations as perceived by users.

Our first future goal is to address current limitations of the techniques: improve the precision of *Touch + Touch*, design state-aware overlays for *Pen + Pen*, support richer audio or vibrotactile feedback to reduce errors. Second, we are planning to evaluate our interface for writing music with expert musicians. This will help us assess its potential but also observe the weaknesses and strengths of the four techniques in more ecological settings.

ACKNOWLEDGMENTS

This work has been partly funded by ANR (ANR-11-JS02-004-01). I thank Jérémie Garcia, Pierre Rossinès, Wendy Mackay, and members of InSitu for fruitful discussions.

REFERENCES

1. Anstice, J., T. Bell, A. Cockburn, and M. Setchell. The Design of a Pen-Based Musical Input System. In Proc. OzCHI '96, 260-267. 1996.
2. Balakrishnan, R., G. Fitzmaurice, G. Kurtenbach, and W. Buxton. Digital Tape Drawing. In Proc. UIST '99, 161-169. ACM, 1999.
3. Balakrishnan, R. and K. Hinckley. The Role of Kinesthetic Reference Frames in Two-Handed Input Performance. In Proc. UIST '99, 171-178. ACM, 1999.
4. Balakrishnan, R. and K. Hinckley. Symmetric Bimanual Interaction. In Proc. CHI '00, 33-40. ACM, 2000.
5. Beaudouin-Lafon, M. Instrumental interaction: an interaction model for designing post-WIMP user interfaces. In Proc. CHI '00, 446-453. ACM, 2000.
6. Bragdon, A., E. Nelson, Y. Li, and K. Hinckley. Experimental analysis of touch-screen gesture designs in mobile environments. In Proc. CHI '11, 403-412. ACM, 2011.
7. Brandl, P., C. Forlines, D. Wigdor, M. Haller, and C. Shen. Combining and Measuring the Benefits of Bimanual Pen and Direct-Touch Interaction on Horizontal Interfaces. In Proc. AVI '08, 154-161. ACM, 2008.
8. Buxton, B. Chunking and phrasing and the design of human-computer dialogues. In Proc. IFIP World Computer Congress, 475-480. 1986.
9. Dillon, R.F., J.D. Edey, and J.W. Tombaugh. Measuring the true cost of command selection: techniques and results. In Proc. CHI '90, 19-26. ACM, 1990.
10. Forsberg, A., M. Dieterich, and R. Zeleznik. The Music Notepad In Proc. UIST '98, 203-210. ACM, 1998.
11. Garcia, J., T. Tsandilas, C. Agon, and W. Mackay. Interactive Paper Substrates to support Musical Creation. In Proc. CHI '12, 1825-1828. ACM, 2012.
12. Guiard, Y., Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. Journal of Motor Behavior, 1987. 19(4): p. 486-517.
13. Heinrichs, F., D. Schreiber, J. Huber, and M. Mühlhäuser. Toward a theory of interaction in mobile paper-digital ensembles. In Proc. CHI '12, 1897-1900. ACM, 2012.
14. Heinrichs, F., J. Steimle, D. Schreiber, and M. Mühlhäuser. Letras: an architecture and framework for ubiquitous pen-and-paper interaction. In Proc. EICS '10, 193-198. ACM, 2010.
15. Hinckley, K., K. Yatani, M. Pahud, N. Coddington, J. Rodenhouse, A. Wilson, H. Benko, and B. Buxton. Pen + touch = new tools. In Proc. UIST '10, 27-36. ACM, 2010.
16. Li, Y., K. Hinckley, Z. Guan, and J.A. Landay. Experimental Analysis of Mode Switching Techniques in Pen-based User Interfaces. In Proc. CHI '05, 461-470. ACM, 2005.
17. Liao, C., F. Guimbretière, K. Hinckley, and J. Hollan. Papiercraft: A gesture-based command system for interactive paper. ACM TOCHI, 2008. 14(4): p. 1-27.
18. Liao, C., Q. Liu, B. Liew, and L. Wilcox. PACER: Fine-grained Interactive Paper via Camera-touch Hybrid Gestures on a Cell Phone. In Proc. CHI '10, 2441-2450. ACM, 2010.
19. Mackay, W.E., G. Pothier, C. Letondal, K. Bøegh, and H.E. Sørensen. The missing link: augmenting biology laboratory notebooks. In Proc. UIST '02, 41-50. ACM, 2002.
20. Rebelo, A., I. Fujinaga, F. Paszkiewicz, A.R.S. Marcal, C. Guedes, and J.S. Cardoso. Optical music recognition: state-of-the-art and open issues. International Journal of Multimedia Information Retrieval, Springer London, 2012: p. 1-18.
21. Rubine, D. Specifying gestures by example. In Proc. SIGGRAPH '91, 329-337. ACM, 1991.
22. Song, H., F. Guimbretière, T. Grossman, and G.W. Fitzmaurice. MouseLight: bimanual interactions on digital paper using a pen and a spatially-aware mobile projector. In Proc. CHI '10, 2451-2460. ACM, 2010.
23. Swinnen, S.P., K. Jardin, and R. Meulenbroek. Between-limb asynchronies during bimanual coordination: Effects of manual dominance and attentional cueing. Neuropsychologia, Elsevier Science, 1996. 34(12): p. 1203-1213.
24. Tsandilas, T., C. Letondal, and W.E. Mackay. Musink: composing music through augmented drawing. In Proc. CHI '09, 819-828. ACM, 2009.
25. Tsandilas, T. and W. Mackay. Knotty Gestures: Subtle Traces to Support Interactive Use of Paper In Proc. AVI '10, 147-154. ACM, 2010.
26. Yeh, R., C. Liao, S. Klemmer, F. Guimbretière, B. Lee, B. Kakaradov, J. Stamberger, and A. Paepcke. ButterflyNet: A Mobile Capture and Access System for Field Biology Research. In Proc. CHI '06, 571-580. ACM, 2006.
27. Zeleznik, R., A. Bragdon, F. Adeptura, and H.-S. Ko. Hands-On Math: A page-based multi-touch and pen desktop for technical work and problem solving. In Proc. UIST '10, 17-26. ACM, 2010.